

# The smallest singular value of deformed random rectangular matrices

Fan Yang <sup>\*1</sup>

<sup>1</sup>Department of Mathematics, University of Wisconsin-Madison

February 15, 2017

## Abstract

We prove an estimate on the smallest singular value of a multiplicatively and additively deformed random rectangular matrix. Suppose  $n \leq N \leq M \leq \Lambda N$  for some constant  $\Lambda \geq 1$ . Let  $X$  be an  $M \times n$  random matrix with independent and identically distributed entries, which have zero mean, unit variance and arbitrarily high moments. Let  $T$  be an  $N \times M$  deterministic matrix with comparable singular values  $c \leq s_N(T) \leq s_1(T) \leq c^{-1}$  for some constant  $c > 0$ . Let  $A$  be an  $N \times n$  deterministic matrix with  $\|A\| = O(\sqrt{N})$ . Then we prove that for any  $\epsilon > 0$ , the smallest singular value of  $TX - A$  is larger than  $N^{-\epsilon}(\sqrt{N} - \sqrt{n-1})$  with high probability. If we assume further the entries of  $X$  have subgaussian decay, then the smallest singular value of  $TX - A$  is at least of the order  $\sqrt{N} - \sqrt{n-1}$  with high probability, which is an essentially optimal estimate.

## 1 Introduction

### 1.1 Smallest singular values of random matrices

Consider an  $N \times n$  real or complex matrix  $A$ . The singular values  $s_i(A)$  of  $A$  are the eigenvalues of  $(A^*A)^{1/2}$  arranged in the non-increasing order:

$$s_1(A) \geq s_2(A) \geq \dots \geq s_n(A).$$

Of particular importance are the largest singular value  $s_1(A)$ , which gives the spectral norm  $\|A\|$ , and the smallest singular value  $s_n(A)$ , which measures the invertibility of  $A^*A$  in the  $N \geq n$  case.

A natural random matrix model is given by a rectangular matrix  $X$  whose entries are independent random variables with mean zero, unit variance and certain moment assumptions. In this paper, we focus on random variables with *arbitrarily high moments* (see (1.6)), which include all the *subgaussian* and *subexponential* random variables. The asymptotic behavior of the extreme singular values of  $X$  has been well-studied. Suppose  $X$  has dimensions  $N \times n$ . Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  be the eigenvalues of  $N^{-1}X^*X$  and define the empirical spectral distribution as  $\mu_N := n^{-1} \sum_{i=1}^n \delta_{\lambda_i}$ . If  $n/N \rightarrow \lambda \in (0, 1)$  as  $N \rightarrow \infty$ , then  $\mu_N$  converges weakly to the famous Marchenko-Pastur (MP) law

---

\*E-mail: fyang75@math.wisc.edu. Partially supported by NSF Career Grant DMS-1552192.

[13]. Moreover, the MP distribution has a density with positive support on  $[(1 - \sqrt{\lambda})^2, (1 + \sqrt{\lambda})^2]$ , which suggests that asymptotically,

$$s_1(X) \rightarrow \sqrt{N}(1 + \sqrt{\lambda}) = \sqrt{N} + \sqrt{n}, \text{ and } s_n(X) \rightarrow \sqrt{N}(1 - \sqrt{\lambda}) = \sqrt{N} - \sqrt{n}. \quad (1.1)$$

The almost sure convergence of the largest singular value was proved in [7] for random matrices whose entries have arbitrarily high moments. The almost sure convergence of the smallest singular value was proved in [20] for Gaussian random matrices (i.e. the Wishart matrix). These results were later generalized to random matrices with *i.i.d.* entries with finite fourth moment in [24] and [3].

A considerably harder problem is to establish non-asymptotic versions of (1.1), which would hold for any fixed dimensions  $N$  and  $n$ . Most often needed are upper bounds for the largest singular value  $s_1(X)$  and lower bounds for the smallest singular value  $s_n(X)$ . With a standard  $\epsilon$ -net argument, it is not hard to prove that  $\|X\|$  is at most of the optimal order  $\sqrt{N}$  for all dimensions, see e.g. [6, 11, 17]. On the other hand, the smallest singular value is much harder to bound below. There has been much progress in this direction during the last decade.

**Tall matrices.** It was proved in [12] that for arbitrary aspect ratios  $\lambda < 1 - c/\log N$  and for random matrices with independent subgaussian entries, one has

$$\mathbb{P}\left(s_n(X) \leq c_\lambda \sqrt{N}\right) \leq e^{-cN}, \quad (1.2)$$

where  $c_\lambda > 0$  depends only on  $\lambda$  and the maximal subgaussian moment of the entries.

**Square matrices.** For square random matrices with  $N = n$ , a lower bound for the smallest singular value was first obtained in [15], where it was proved that for subgaussian random matrix  $X$ ,  $s_N(X) \geq \epsilon N^{-3/2}$  with high probability. This result was later improved in [16] to

$$\mathbb{P}\left(s_N(X) \leq \epsilon N^{-1/2}\right) \leq C\epsilon + e^{-cN}, \quad (1.3)$$

an essentially optimal estimate for subgaussian matrices. Subsequently, different lower bounds for  $s_N(X)$  were proved under weakened moments assumptions [21, 14, 9].

**Almost square matrices.** The gap  $1 - c/\log N \leq \lambda < 1$  was filled in [17]. It was shown that for subgaussian random rectangular matrices,

$$\mathbb{P}\left(s_n(X) \leq \epsilon(\sqrt{N} - \sqrt{n-1})\right) \leq (C\epsilon)^{N-n+1} + e^{-cN}, \quad (1.4)$$

for all fixed dimensions  $N \geq n$ . This bound is essentially optimal for subgaussian matrices with all aspect ratios. It is easy to see that (1.2) and (1.3) are the special cases of the estimate (1.4).

In this paper, we are interested in the extreme singular values of a multiplicatively and additively deformed random rectangular matrix. Given an  $M \times n$  random matrix  $X$  with independent entries, we consider the matrix  $TX - A$ , where  $T$  and  $A$  are  $N \times M$  and  $N \times n$  deterministic matrices, respectively. It is easy to bound above the largest singular value using  $\|TX - A\| \leq \|T\|\|X\| + \|A\|$ . On the other hand, we expect that if  $n \leq N \leq M$  and the singular values of  $T$  satisfy  $c \leq s_N(T) \leq s_1(T) \leq c^{-1}$ , then a similar estimate as in (1.4) would still hold for  $TX - A$ . In fact, if  $M = N$  and  $X$  is subgaussian, one can prove that the estimate (1.4) holds for the matrix  $X - T^{-1}A$  with a direct generalization of the method in [17]. Together with  $s_n(TX - A) \geq s_N(T)s_n(X - T^{-1}A)$ , this already gives the desired lower bound for  $s_n(TX - A)$ . In this paper, we will consider more general case where  $N \leq M$  and  $X$  is not necessarily subgaussian, see Theorem 1.1.

One of our motivations is the potential application in statistical science. Consider sample covariance matrices of the form  $Q = n^{-1}BB^*$ , where  $B$  is an  $N \times n$  matrix. The columns of  $B$  represent  $n$  independent observations of some random  $N$ -dimensional vector  $\mathbf{b}$ . For the sample vector  $\mathbf{b}$ , we take

a linear model  $\mathbf{b} = T\mathbf{x}$ , where  $T$  is a deterministic  $N \times M$  matrix and  $\mathbf{x}$  is a random  $M$ -dimensional vector with independent entries. Then we can write  $\mathbf{b} = T\hat{\mathbf{x}} + \mathbf{a}$ , where  $\hat{\mathbf{x}}$  is a centered random vector and  $\mathbf{a} = T\mathbb{E}\mathbf{x}$ . In addition, without loss of generality, we may assume that the entries of  $\hat{\mathbf{x}}$  have unit variance by absorbing the variance of  $\hat{x}_i$  into  $T$ . Hence we can write  $B$  into the form  $B = TX - A$ , and our result would provide a good a priori estimate on the smallest singular values of  $B$ .

Another application of our result is the *circular law* for square random matrices. Let  $X$  be an  $N \times N$  random matrix with *i.i.d.* entries with zero mean and unit variance. It is well known that the spectral measure of eigenvalues of  $N^{-1/2}X$  converges to the circular law, i.e. the uniform distribution on the unit disk [8, 1]. An important input of the proof is the lower bound for the smallest singular value of  $N^{-1/2}X - z$  for any  $z \in \mathbb{C}$  [9, 14, 21, 22]. In [23], we proved a generalized *local circular law* for square random matrices of the form  $N^{-1/2}TX$ . In order to obtain a lower bound for the smallest singular value of  $N^{-1/2}TX - z$ , we assumed that the entries of  $X$  have continuous distributions. This assumption rules out some important models such as the Bernoulli random matrices. Now with the result of this paper, we can relax this assumption greatly to include all random variables with sufficiently high moments.

## 1.2 Main result and the reduction to subgaussian matrices

Let  $\xi_1, \dots, \xi_n$  be independent random variables such that for  $1 \leq i \leq n$ ,

$$\mathbb{E}\xi_i = 0, \quad \mathbb{E}|\xi_i|^2 = 1, \quad (1.5)$$

and for any  $p \in \mathbb{N}$ , there is an  $N$ -independent constant  $\sigma_p$  such that

$$\mathbb{E}|\xi_i|^p \leq \sigma_p. \quad (1.6)$$

We assume that  $X$  is an  $M \times n$  random matrix, whose rows are independent copies of the random vector  $(\xi_1, \dots, \xi_n)$ . In this paper, we consider the deformed random rectangular matrix  $TX - B$ , where  $T$  and  $B$  are  $N \times M$  and  $N \times n$  deterministic matrices, respectively. We assume that

$$n \leq N \leq M \leq \Lambda N, \quad \|B\| \leq K_0\sqrt{N} \quad (1.7)$$

for some constants  $K_0, \Lambda \geq 1$ . Moreover, we assume the eigenvalues of  $TT^*$  satisfy that

$$K_0^{-1} \leq \sigma_N \leq \dots \leq \sigma_2 \leq \sigma_1 \leq K_0. \quad (1.8)$$

For definiteness, in this paper we focus on the case with *real* matrices. However, our results and proof also hold, after minor changes, in the *complex* case if we assume in addition that  $X_{ij}$  have independent real and imaginary parts, such that

$$\mathbb{E}(\operatorname{Re} X_{ij}) = 0, \quad \mathbb{E}(\operatorname{Re} X_{ij})^2 = \frac{1}{2},$$

and similarly for  $\operatorname{Im} X_{ij}$ . The main result of this paper is the following theorem.

**Theorem 1.1.** *Suppose the assumptions (1.5), (1.6), (1.7) and (1.8) hold. Fix any constants  $\tau > 0$  and  $\Gamma > 0$ . Then for every  $\epsilon \geq 0$ , we have*

$$\mathbb{P}\left(s_n(TX - B) \leq \epsilon N^{-\tau} \left(\sqrt{N} - \sqrt{n-1}\right)\right) \leq (C\epsilon)^{N-n+1} + N^{-\Gamma} \quad (1.9)$$

for large enough  $N \geq N_0$ , where the constant  $C > 0$  depends only on  $\sigma_p$ ,  $\Lambda$  and  $K_0$ , and  $N_0$  depends only on  $\sigma_p$ ,  $\Lambda$ ,  $\Gamma$  and  $\tau$ .

To prove this theorem, we first truncate the entries of  $X$  at level  $N^\omega$  for some small  $\omega > 0$ . Combining condition (1.6) with Markov's inequality, we get that for any (small)  $\omega > 0$  and (large)  $\Gamma > 0$ , there exists  $N(\omega, \Gamma)$  such that

$$\mathbb{P}(|\xi_i| > N^\omega/2) \leq N^{-\Gamma-2}$$

for all  $N \geq N(\omega, \Gamma)$ . Hence with a loss of probability  $O(N^{-\Gamma})$ , it suffices to control the smallest singular values of the random matrix  $T\tilde{X} - B$ , where

$$\tilde{X} := 1_\Omega X, \quad \Omega := \{|X_{ij}| \leq N^\omega/2 \text{ for all } 1 \leq i \leq M, 1 \leq j \leq n\}.$$

By (1.6) and integration by parts, we can check that for  $1 \leq i \leq n$ ,

$$\mathbb{E}(\xi_i \mathbf{1}_{\{|\xi_i| \leq N^\omega/2\}}) = O(N^{-\Gamma-2+\omega}), \quad \text{Var}(\xi_i \mathbf{1}_{\{|\xi_i| \leq N^\omega/2\}}) = 1 + O(N^{-\Gamma-2+2\omega}). \quad (1.10)$$

We define  $D_1$  to be an  $n \times n$  diagonal matrix with  $(D_1)_{ii} = \text{Var}(\xi_i \mathbf{1}_{\{|\xi_i| \leq N^\omega/2\}})^{1/2}$ .

Let  $T = U\tilde{D}V$  be a singular value decomposition of  $T$ , where  $U$  is an  $N \times N$  unitary matrix,  $V$  is an  $M \times M$  unitary matrix and  $\tilde{D} = (D, 0)$  is an  $N \times M$  rectangular diagonal matrix such that  $D = \text{diag}(d_1, d_2, \dots, d_N)$  with  $d_i^2 = \sigma_i$ . We denote  $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$ , where  $V_1$  has size  $N \times M$  and  $V_2$  has size  $(M - N) \times M$ . Then we have

$$T\tilde{X} - B = UDV_1(\tilde{X} - \mathbb{E}\tilde{X}) - (B - T\mathbb{E}\tilde{X}) = UD \left[ V_1(\tilde{X} - \mathbb{E}\tilde{X})D_1^{-1} - (D^{-1}U^{-1}B - V_1\mathbb{E}\tilde{X})D_1^{-1} \right] D_1.$$

Due to (1.8) and (1.10), we only need to bound  $s_n(V_1Y - A)$ , where

$$Y := (\tilde{X} - \mathbb{E}\tilde{X})D_1^{-1}, \quad \text{and} \quad A := (D^{-1}U^{-1}B - V_1\mathbb{E}\tilde{X})D_1^{-1}.$$

Using (1.7), (1.8), (1.10) and the definition of  $\Omega$ , it is easy to check that  $A$  is a deterministic matrix with

$$\|A\| \leq C(\|B\| + \|\mathbb{E}\tilde{X}\|) \leq C(\sqrt{N} + N^{-\Gamma-1+\omega}) \leq C\sqrt{N}, \quad (1.11)$$

and  $Y$  is a random matrix with independent entries satisfying

$$\mathbb{E}(Y_{ij}) = 0, \quad \text{Var}(Y_{ij}) = 1, \quad |Y_{ij}| \leq N^\omega. \quad (1.12)$$

Recall that a random variable  $\xi$  is called subgaussian if there exists  $K > 0$  such that

$$\mathbb{P}(|\xi| > t) \leq 2 \exp(-t^2/K^2) \quad \text{for all } t > 0. \quad (1.13)$$

The infimum of such  $K$  is called the subgaussian moment of  $\xi$  or the  $\psi_2$ -norm  $\|\xi\|_{\psi_2}$ . By (1.12), it is obvious that  $Y_{ij}$  are subgaussian random variables with  $\|Y_{ij}\|_{\psi_2} \leq N^\omega$ . Moreover, by Theorem 2.10 of [5], there exists a constant  $C > 0$  such that

$$\mathbb{P}(\|X\| \leq C\sqrt{N}) \geq 1 - N^{-\Gamma}$$

for large enough  $N$ . Then using  $\|\tilde{X}\| \leq \|X\|$ , we get that

$$\mathbb{P}(\|Y\| \leq C\sqrt{N}) \geq 1 - N^{-\Gamma}. \quad (1.14)$$

From the above discussion, we see that Theorem 1.1 follows from the following theorem.

**Theorem 1.2.** Let  $\xi_1, \dots, \xi_n$  be independent centered random variables with unit variance, finite fourth moments and subgaussian moments bounded by  $K$  for some  $K \equiv K(N) \leq N^\omega$ . Let  $Y$  be an  $M \times n$  random matrix, whose rows are independent copies of the random vector  $(\xi_1, \dots, \xi_n)$ . Let  $P$  be an  $N \times M$  deterministic matrix with  $PP^T = 1$ , and let  $A$  be an  $N \times n$  deterministic matrix. Suppose that  $\|Y\| + \|A\| \leq C_1\sqrt{N}$  for some constant  $C_1 > 0$ . Then for every  $0 < \omega < \omega_0$  and every  $\epsilon \geq 0$ , we have

$$\mathbb{P}\left(s_n(PY - A) \leq \epsilon \left(\sqrt{N} - \sqrt{n-1}\right)\right) \leq (CK^L\epsilon)^{N-n+1} + e^{-cN/K^4}, \quad (1.15)$$

where the constants  $\omega_0, c, C, L > 0$  depend only on  $\Lambda, C_1$  and the maximal fourth moment.

**Remark 1.3.** Suppose  $X_{ij}$  are subgaussian random variables with  $\max_{i,j} \|X_{ij}\| \leq K$  for some constant  $K > 0$ . Then we have

$$\mathbb{P}(\|X\| \geq t\sqrt{N}) \leq e^{-c_0 t^2 N} \quad \text{for } t \geq C_0,$$

where  $c_0, C_0 > 0$  depend only on  $K$  (see [17, Proposition 2.4]). Combining with Theorem 1.2, we obtain the optimal estimate for the smallest singular value of  $TX - B$ :

$$\mathbb{P}\left(s_n(TX - B) \leq \epsilon \left(\sqrt{N} - \sqrt{n-1}\right)\right) \leq (C\epsilon)^{N-n+1} + e^{-cN}. \quad (1.16)$$

The bulk of this paper is devoted to the proof of Theorem 1.2. In the preliminary Section 2, we introduce some notations and tools that will be used in the proof. In Section 3, we first reduce the problem into bounding below  $\|(PY - A)x\|_2$  for *compressible unit vectors*  $x \in S^{n-1}$ , whose  $l^2$ -norm is concentrated in a small number of coordinates, and for *incompressible unit vectors* comprising the rest of the sphere  $S^{n-1}$ . Then we prove a lower bound for compressible unit vectors using a small ball probability result (Lemma 2.7) and a standard  $\epsilon$ -net argument. The incompressible unit vectors are dealt with in Sections 4 and 5. In Section 4, we consider the case  $1 \leq n \leq \lambda_0 N$  for some constant  $\lambda_0 \in (0, 1)$ , i.e. when  $PY - A$  is a tall matrix. The proof can be finished with another small ball probability result (Lemma 2.6) and the  $\epsilon$ -net argument. The almost square case with  $\lambda_0 N < n \leq N$  is considered in Section 5. We first reduce the problem into bounding the distance between a random vector and a random subspace, and then complete the proof with a random distance lemma—Lemma 5.3, whose proof will be given in Section 6.

**Acknowledgements.** I would like to thank Roman Vershynin for valuable suggestions and for pointing out several useful references. I am also grateful to my advisor Jun Yin for his financial support and helpful comments on this paper. Finally, many thanks to my friend Haokai Xi for useful discussions.

## 2 Basic notations and tools

In this paper, we use  $C$  to denote a generic large positive constant, which may depend on fixed parameters and whose value may change from one line to the next. Similarly, we use  $c, \epsilon$  or  $\omega$  to denote a generic small positive constant. If a constant depends on a quantity  $a$ , we use  $C(a)$  or  $C_a$  to indicate this dependence.

The canonical inner product on  $\mathbb{R}^n$  is denoted  $\langle \cdot, \cdot \rangle$ , and the Euclidean norm is denoted  $\|\cdot\|_2$ . The distance from a point  $x$  to a set  $D$  in  $\mathbb{R}^n$  is denoted  $\text{dist}(x, D)$ . The unit sphere centered at the origin in  $\mathbb{R}^n$  is denoted  $S^{n-1}$ . The orthogonal projection in  $\mathbb{R}^n$  onto a subspace  $E$  is denoted

$P_E$ . For a subset of coordinates  $J \subseteq \{1, \dots, n\}$ , we often write  $P_J$  for  $P_{\mathbb{R}^J}$ . The unit sphere of  $E$  is denoted  $S(E) := S^{n-1} \cap E$ .

For any matrix  $A$ , we use  $A^*$  to denote its conjugate transpose,  $A^T$  the transpose,  $\|A\| := \|A\|_{l^2 \rightarrow l^2}$  the operator norm and  $\|A\|_{HS}$  the Hilbert-Schmidt norm. We usually write an identity matrix as  $I$  without causing any confusions.

The following tensorization lemma is Lemma 2.2 of [16]

**Lemma 2.1** (Tensorization). *Let  $\zeta_1, \dots, \zeta_n$  be independent non-negative random variables, and let  $B, \epsilon_0 \geq 0$ .*

(1) *Assume that for each  $k$ ,*

$$\mathbb{P}(\zeta_k < \epsilon) \leq B\epsilon \quad \text{for all } \epsilon \geq \epsilon_0.$$

*Then*

$$\mathbb{P}\left(\sum_{k=1}^n \zeta_k^2 < \epsilon^2 n\right) \leq (CB\epsilon)^n \quad \text{for all } \epsilon \geq \epsilon_0,$$

*where  $C$  is an absolute constant.*

(2) *Assume that there exist  $\lambda > 0$  and  $\mu \in (0, 1)$  such that for each  $k$ ,*

$$\mathbb{P}(\zeta_k < \lambda) \leq \mu.$$

*Then there exists  $\lambda_1 > 0$  and  $\mu_1 \in (0, 1)$  that depend on  $\lambda$  and  $\mu$  only and such that*

$$\mathbb{P}\left(\sum_{k=1}^n \zeta_k^2 < \lambda_1 n\right) \leq \mu_1^n.$$

Consider a subset  $\Omega \subset \mathbb{R}^n$ , and let  $\epsilon > 0$ . An  $\epsilon$ -net of  $\Omega$  is a subset  $\mathcal{N} \subseteq \Omega$  such that for every  $x \in \Omega$  one has  $\text{dist}(x, \mathcal{N}) \leq \epsilon$ . The following lemma is proved as Propositions 2.1 and 2.2 in [17].

**Lemma 2.2** (Nets). *Fix any  $\epsilon > 0$ .*

(1) *There exists an  $\epsilon$ -net of  $S^{n-1}$  of cardinality at most*

$$\min \left\{ (1 + 2\epsilon^{-1})^n, 2n (1 + 2\epsilon^{-1})^{n-1} \right\}.$$

(2) *Let  $S$  be a subset of  $S^{n-1}$ . There exists an  $\epsilon$ -net of  $S$  of cardinality at most*

$$\min \left\{ (1 + 4\epsilon^{-1})^n, 2n (1 + 4\epsilon^{-1})^{n-1} \right\}.$$

Next we define the small ball probability for a random vector.

**Definition 2.3.** *The Lévy concentration function of a random vector  $S \in \mathbb{R}^m$  is defined for  $\epsilon > 0$  as*

$$\mathcal{L}(S, \epsilon) = \sup_{v \in \mathbb{R}^m} \mathbb{P}(\|S - v\|_2 \leq \epsilon),$$

*which measures the small ball probabilities.*

With Definition 2.3, it is easy to prove the following lemma. It will allow us to select a nice subset of the coefficients  $a_k$  when computing the small ball probability.

**Lemma 2.4.** Let  $\xi_1, \dots, \xi_n$  be independent random variables. For any  $\sigma \subseteq \{1, \dots, n\}$ , any  $a \in \mathbb{R}^n$  and any  $\epsilon \geq 0$ , we have

$$\mathcal{L}\left(\sum_{k=1}^n a_k \xi_k, \epsilon\right) \leq \mathcal{L}\left(\sum_{k \in \sigma} a_k \xi_k, \epsilon\right).$$

The following three lemmas give some useful small ball probability bounds. They correspond to [17, Lemma 3.2], [16, Corollary 2.9] and [18, Corollary 2.4] respectively.

**Lemma 2.5.** Let  $\xi$  be a random variable with mean zero, unit variance, and finite fourth moment. Then for every  $\epsilon \in (0, 1)$ , there exists a  $p \in (0, 1)$  which depends only on  $\epsilon$  and on the fourth moment, and such that

$$\mathcal{L}(\xi, \epsilon) \leq p.$$

**Lemma 2.6.** Let  $\xi_1, \dots, \xi_n$  be independent centered random variables with variances at least 1 and third moments bounded by  $B$ . Then for every  $a \in \mathbb{R}^n$  and every  $\epsilon \geq 0$ , one has

$$\mathcal{L}\left(\sum_{k=1}^n a_k \xi_k, \epsilon\right) \leq \sqrt{\frac{2}{\pi}} \frac{\epsilon}{\|a\|_2} + \tilde{C} B \left(\frac{\|a\|_3}{\|a\|_2}\right)^3,$$

where  $\tilde{C}$  is an absolute constant.

**Lemma 2.7.** Let  $A$  be a fixed  $N \times M$  matrix. Consider a random vector  $\xi = (\xi_1, \dots, \xi_M)$  where  $\xi_i$  are independent random variables satisfying  $\mathbb{E}\xi_i = 0$ ,  $\mathbb{E}\xi_i^2 = 1$  and  $\|\xi_i\|_{\psi_2} \leq K$ . Then for every  $y \in \mathbb{R}^N$ , we have

$$\mathbb{P}\left\{\|A\xi - y\|_2 \leq \frac{1}{2}\|A\|_{HS}\right\} \leq 2 \exp\left(-\frac{c\|A\|_{HS}^2}{K^4\|A\|^2}\right).$$

### 3 Decomposition of the sphere

Now we begin the proof of Theorem 1.2. We will make use of a partition of the unit sphere into two sets of compressible and incompressible vectors. They are first defined in [16].

**Definition 3.1.** Let  $\delta, \rho \in (0, 1]$ . A vector  $x \in \mathbb{R}^n$  is called *sparse* if  $|\text{supp}(x)| \leq \delta n$ . A vector  $x \in S^{n-1}$  is called *compressible* if  $x$  is within Euclidean distance  $\rho$  from the set of all sparse vectors. A vector  $x \in S^{n-1}$  is called *incompressible* if it is not compressible. The sets of sparse, compressible and incompressible vectors will be denoted by  $\text{Sparse}_n(\delta)$ ,  $\text{Comp}_n(\delta, \rho)$  and  $\text{Incomp}_n(\delta, \rho)$ . We sometimes omit the subindex  $n$  when the dimension is clear.

Using the decomposition  $S^{n-1} = \text{Comp} \cup \text{Incomp}$ , we break the invertibility problem into two subproblems, for compressible and incompressible vectors:

$$\mathbb{P}\left(s_n(PY - A) \leq \epsilon(\sqrt{N} - \sqrt{n-1})\right) \leq \mathbb{P}\left(\inf_{x \in \text{Comp}_n(\delta, \rho)} \|(PY - A)x\|_2 \leq \epsilon(\sqrt{N} - \sqrt{n-1})\right) \quad (3.1)$$

$$+ \mathbb{P}\left(\inf_{x \in \text{Incomp}_n(\delta, \rho)} \|(PY - A)x\|_2 \leq \epsilon(\sqrt{N} - \sqrt{n-1})\right). \quad (3.2)$$

The bound for compressible vectors follows from the following lemma, which is a variant of Lemma 3.3 from [16].

**Lemma 3.2.** *Suppose the assumptions in Theorem 1.2 hold. Then there exist  $\rho, c_0, c_1 > 0$  that depend only on  $C_1$ , and such that for  $\delta \leq \min\{c_1 N / (nK^4 \log K), 1\}$ , we have*

$$\mathbb{P} \left( \inf_{x \in \text{Comp}_n(\delta, \rho)} \|(PY - A)x\|_2 \leq c_0 \sqrt{N} \right) \leq e^{-c_0 N / K^4}.$$

*Proof.* We first prove a similar estimate for sparse vectors. For any  $x \in S^{n-1}$ , we define the random vector  $\zeta := Yx \in \mathbb{R}^N$ . It is easy to verify that  $\mathbb{E}\zeta_i = 0$ ,  $\mathbb{E}\zeta_i^2 = 1$  and  $\|\zeta_i\|_{\psi_2} \leq CK$ . Then with  $\|P\| = 1$  and  $\|P\|_{HS}^2 = N$ , we conclude from Lemma 2.7 that

$$\mathbb{P} \left\{ \|(PY - A)x\|_2 \leq \frac{1}{2} \sqrt{N} \right\} \leq 2 \exp \left( -\frac{cN}{K^4} \right). \quad (3.3)$$

Let  $S_1 := \{x \in S^{n-1} : x_k = 0, k > \lceil \delta n \rceil\}$ . By Lemma 2.2, there exists an  $\epsilon$ -net  $\mathcal{N}$  of  $S_1$  with  $|\mathcal{N}| \leq (5/\epsilon)^{\lceil \delta n \rceil}$ . Then using (3.3) and taking the union bound, we get

$$\mathbb{P} \left( \inf_{x \in \mathcal{N}} \|(PY - A)x\|_2 \leq \frac{1}{2} \sqrt{N} \right) \leq 2e^{-cN/K^4} (5\epsilon^{-1})^{\lceil \delta n \rceil}. \quad (3.4)$$

Let  $V$  be the event that  $\|(PY - A)y\| \leq \sqrt{N}/4$  for some  $y \in S_1$ . By the assumptions of Theorem 1.2, we have

$$\|PY - A\| \leq \|Y\| + \|A\| \leq C_1 \sqrt{N}.$$

Assume that  $V$  occurs and choose a point  $x \in \mathcal{N}$  such that  $\|y - x\| \leq \epsilon$ . Then

$$\|(PY - A)x\|_2 \leq \|(PY - A)y\|_2 + \|PY - A\| \|x - y\|_2 \leq \frac{1}{4} \sqrt{N} + C_1 \epsilon \sqrt{N} \leq \frac{1}{2} \sqrt{N},$$

if we choose  $\epsilon \leq 1/(4C_1)$ . Fix one such  $\epsilon$ , using (3.4) we obtain that

$$\mathbb{P} \left( \inf_{x \in S_1} \|(PY - A)x\|_2 \leq \frac{1}{4} \sqrt{N} \right) = \mathbb{P}(V) \leq 2e^{-cN/K^4} (5\epsilon^{-1})^{\lceil \delta n \rceil} \leq e^{-c_2 N / K^4},$$

if we choose  $c_1$  (and hence  $\delta$ ) to be sufficiently small. We use this result and take the union bound over all  $\lceil \delta n \rceil$ -element subsets  $\sigma$  of  $\{1, \dots, n\}$ :

$$\begin{aligned} & \mathbb{P} \left( \inf_{x \in \text{Sparse}(\delta) \cap S^{n-1}} \|(PY - A)x\|_2 \leq \frac{1}{4} \sqrt{N} \right) \\ &= \mathbb{P} \left( \exists \sigma, |\sigma| = \lceil \delta n \rceil : \inf_{x \in \mathbb{R}^\sigma \cap S^{n-1}} \|(PY - A)x\|_2 \leq \frac{1}{4} \sqrt{N} \right) \\ &\leq \binom{n}{\lceil \delta n \rceil} e^{-c_2 N / K^4} \leq \exp \left( 4e\delta \log \left( \frac{e}{\delta} \right) n - \frac{c_2 N}{K^4} \right) \leq \exp \left( -\frac{c_2 N}{2K^4} \right), \end{aligned} \quad (3.5)$$

with an appropriate choice of  $c_1$ .

Now we deduce the estimate for compressible vectors. Let  $c_3 \geq 0$  and  $\rho \in (0, 1/2)$  to be chosen later. We need to control the event  $W$  that  $\|(PY - A)x\|_2 \leq c_3 \sqrt{N}$  for some vector  $x \in \text{Comp}(\delta, \rho)$ . Assume  $W$  occurs, then every such vector  $x$  can be written as a sum  $x = y + z$  with  $y \in \text{Sparse}(\delta)$  and  $\|z\|_2 \leq \rho$ . Thus  $\|y\|_2 \geq 1 - \rho \geq 1/2$ , and

$$\|(PY - A)y\|_2 \leq \|(PY - A)x\|_2 + \|(PY - A)\| \|z\|_2 \leq c_3 \sqrt{N} + \rho C_1 \sqrt{N}.$$

We choose  $c_3 = 1/16$  and  $\rho = 1/(16C_1)$ , so that  $\|(PY - A)y\|_2 \leq \sqrt{N}/8$ . Since  $\|y\|_2 \geq 1/2$ , we can find a unit vector  $u = y/\|y\|_2 \in \text{Sparse}(\delta)$  such that  $\|(PY - A)u\|_2 \leq \sqrt{N}/4$ . This shows that event  $W$  implies the event in (3.5), so we have  $\mathbb{P}(W) \leq e^{-c_2 N / (2K^4)}$ . This concludes the proof.  $\square$



**Remark 3.3.** If  $n < c_1 N / (K^4 \log K)$ , then all the vectors in  $S^{n-1}$  are in  $\text{Comp}(\delta, \rho)$  and Lemma 3.2 already concludes the proof of Theorem 1.2. Hence throughout the following sections, it suffices to assume

$$n \geq c_1 N / (K^4 \log K). \quad (3.6)$$

It remains to prove the bound for incompressible vectors in (3.2). Define the aspect ratio  $\lambda := n/N$ . We will divide the proof into two cases: the case where  $c_1 / (K^4 \log K) \leq \lambda \leq \lambda_0$  for some constant  $0 < \lambda_0 < 1$ , and the case where  $\lambda_0 < \lambda \leq 1$ . We record here an important property of the incompressible vectors, which is proved in Lemma 3.4 of [16].

**Lemma 3.4** (Incompressible vectors are spread). *Let  $x \in \text{Incomp}_n(\delta, \rho)$ . Then there exists a set  $\sigma \equiv \sigma(x) \subseteq \{1, \dots, n\}$  of cardinality  $|\sigma| \geq \frac{1}{2} \rho^2 \delta n$  and such that*

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}} \quad \text{for all } k \in \sigma. \quad (3.7)$$

## 4 Tall matrices

In this section, we deal with the probability in (3.2) when  $c_1 / (K^4 \log K) \leq \lambda \leq \lambda_0$  for some constant  $\lambda_0 \in (0, 1)$ . The value of  $\lambda_0$  will be chosen later in Section 5 (see (5.8)), and it only depends on  $\Lambda$ ,  $C_1$  and the maximal fourth moment of the entries of  $Y$ . Then it is equivalent to control the probability

$$\mathbb{P} \left( \inf_{x \in \text{Incomp}_n(\delta, \rho)} \|(PY - A)x\|_2 \leq t\sqrt{N} \right)$$

for any  $t \geq 0$ .

Let  $x$  be a vector in  $\text{Incomp}_n(\delta, \rho)$ , where we fix  $\delta = c_1 N / (n K^4 \log K)$  and  $0 < \rho \leq 1/(16C_1)$  (see Lemma 3.2). Take the set  $\sigma$  given by Lemma 3.4. Note that the entries of  $Yx$  are of the form  $(Yx)_i = \sum_{k=1}^n Y_{ik} x_k$ ,  $1 \leq i \leq M$ , where  $Y_{ik}$  are independent centered random variables with unit variance and bounded fourth moment. Hence we can use Lemma 2.4 and Lemma 2.6 to get that

$$\mathcal{L}((Yx)_i, t) \leq \sqrt{\frac{2}{\pi}} \frac{t}{\|P_\sigma x\|_2} + C \left( \frac{\|P_\sigma x\|_3}{\|P_\sigma x\|_2} \right)^3 \leq C_2(\rho) \left( \frac{t}{\sqrt{\delta}} + \frac{1}{\delta\sqrt{n}} \right), \quad (4.1)$$

for some constant  $C_2(\rho) > 0$  depending only on  $\rho$  and the maximal fourth moment. Here we used the bound

$$\|P_\sigma x\|_2 \geq \frac{1}{2} \rho^2 \sqrt{\delta}, \quad \left( \frac{\|P_\sigma x\|_3}{\|P_\sigma x\|_2} \right)^3 \leq \frac{2}{\rho^2 \delta \sqrt{n}},$$

deduced from Lemma 3.4. With (4.1) as the input, the next lemma provides a small ball probability bound for the random vector  $PYx$ .

**Lemma 4.1** (Corollary 1.4 of [19]). *Consider a random vector  $X = (\xi_1, \dots, \xi_M)$  where  $\xi_i$  are real-valued independent random variables. Let  $t, p \geq 0$  be such that*

$$\mathcal{L}(\xi_i, t) \leq p \quad \text{for all } i = 1, \dots, M.$$

*Let  $P$  be an orthogonal projection in  $\mathbb{R}^M$  onto an  $N$ -dimensional subspace. Then*

$$\mathcal{L}(PX, t\sqrt{N}) \leq (Cp)^N,$$

*where  $C$  is an absolute constant.*

Applying the above lemma to random vector  $Yx$ , we obtain that

$$\mathbb{P}\left(\|(PY - A)x\|_2 \leq t\sqrt{N}\right) \leq \mathcal{L}\left(PYx, t\sqrt{N}\right) \leq \left[C_3\left(\frac{t}{\sqrt{\delta}} + \frac{1}{\delta\sqrt{n}}\right)\right]^N \quad (4.2)$$

for some constant  $C_3 > 0$ . Now we can take a union bound over all  $x$  in an  $\epsilon$ -net of  $Incomp_n(\delta, \rho)$  and complete the proof by approximation.

We first assume that  $t \geq 1/\sqrt{\delta n}$ . Then the  $t/\sqrt{\delta}$  term in (4.2) dominates and we obtain that

$$\mathbb{P}\left(\|(PY - A)x\|_2 \leq t\sqrt{N}\right) \leq \left(2C_3 t/\sqrt{\delta}\right)^N.$$

By Lemma 2.2, there exists an  $\epsilon$ -net  $\mathcal{N}$  in  $Incomp_n(\delta, \rho)$  of cardinality  $|\mathcal{N}| \leq 2n(5/\epsilon)^{n-1}$ . Taking the union bound, we get

$$\mathbb{P}\left(\inf_{x \in \mathcal{N}} \|(PY - A)x\|_2 \leq t\sqrt{N}\right) \leq 2n \left(\frac{2C_3 t}{\sqrt{\delta}}\right)^N \left(\frac{5}{\epsilon}\right)^{n-1}. \quad (4.3)$$

Let  $V$  be the event that  $\|(PY - A)y\|_2 \leq t\sqrt{N}/2$  for some  $y \in Incomp_n(\delta, \rho)$ . Assume that  $V$  occurs and choose a point  $x \in \mathcal{N}$  such that  $\|x - y\|_2 \leq \epsilon$ . Then if  $\epsilon \leq t/(2C_1)$ , we have

$$\|(PY - A)x\|_2 \leq \|(PY - A)y\|_2 + \|PY - A\| \|x - y\|_2 \leq \frac{1}{2}t\sqrt{N} + C_1\epsilon\sqrt{N} \leq t\sqrt{N},$$

where we used that  $\|PY - A\| \leq C_1\sqrt{N}$ . Fix one such  $\epsilon$ , using (4.3) we obtain that

$$\begin{aligned} \mathbb{P}\left(\inf_{x \in Incomp_n(\delta, \rho)} \|(PY - A)x\|_2 \leq \frac{t}{2}\sqrt{N}\right) &= \mathbb{P}(V) \\ &\leq 2n \left(\frac{2C_3 t}{\sqrt{\delta}}\right)^N \left(\frac{10C_1}{t}\right)^{n-1} \leq \left[\left(C_4\delta^{-1/2}\right)^{1/(1-\lambda_0)} t\right]^{N-n+1}, \end{aligned} \quad (4.4)$$

where in the last step we used  $n/N \leq \lambda_0$ . If  $t \leq 1/\sqrt{\delta n}$ , we use (4.4) to get

$$\mathbb{P}\left(\inf_{x \in Incomp_n(\delta, \rho)} \|(PY - A)x\|_2 \leq \frac{t}{2}\sqrt{N}\right) \leq \left[\left(C_4\delta^{-1/2}\right)^{1/(1-\lambda_0)} (\delta n)^{-1/2}\right]^{N-n+1} \leq e^{-c_0 N/K^4},$$

if  $K \leq N^\omega$  for some sufficiently small  $\omega$ . Together with (4.4) and Lemma 3.2, this concludes the proof of Theorem 1.2 for the  $\lambda \leq \lambda_0$  case.

## 5 Almost square matrices

In this section, we deal with the probability in (3.2) for the  $\lambda_0 < \lambda \leq 1$  case. In particular, when  $\lambda \rightarrow 1$ ,  $PY - A$  becomes an almost square matrix and (4.4) cannot provide a satisfactory probability bound. For instance, for the square case with  $N = n$ , it is easy to see that the  $(C\delta^{-1/2})^N$  term dominates over the  $t$  term. To handle this difficulty, we will use the method in [17], which reduces the problem of bounding  $\|(PY - A)x\|_2$  for  $x \in Incomp_n(\delta, \rho)$  to a random distance problem. We denote  $N = n - 1 + d$  for some  $d \geq 1$ . Note that  $\sqrt{N} - \sqrt{n-1} \leq d/\sqrt{n}$ . Hence to bound (3.2), it suffices to bound

$$\mathbb{P}\left(\inf_{x \in Incomp_n(\delta, \rho)} \|(PY - A)x\|_2 \leq \epsilon \frac{d}{\sqrt{n}}\right), \quad \text{for } \delta = c_1 N / (nK^4 \log K), \quad \rho \leq 1/(16C_1). \quad (5.1)$$

We denote

$$m := \min \left\{ d, \left\lfloor \frac{1}{2} \rho^2 \delta n \right\rfloor \right\}. \quad (5.2)$$

Let  $Z_1 := PY_1 - A_1, \dots, Z_n := PY_n - A_n$  be the columns of the matrix  $Z := PY - A$ . Given a subset  $J \subset \{1, \dots, n\}$  of cardinality  $m$ , we define the subspace

$$H_{J^c} := \text{span}(Z_k)_{k \in J^c} \subseteq \mathbb{R}^N. \quad (5.3)$$

For levels  $K_1 := \rho\sqrt{\delta/2}$  and  $K_2 := K_1^{-1}$ , we define the set of totally spread vectors

$$S^J := \left\{ y \in S^{n-1} \cap \mathbb{R}^J : \frac{K_1}{\sqrt{m}} \leq |y_k| \leq \frac{K_2}{\sqrt{m}} \text{ for all } k \in J \right\}. \quad (5.4)$$

In the following lemma, we let  $J$  be a random subset uniformly chosen over all subsets of  $\{1, \dots, n\}$  of cardinality  $m$ . We shall write  $P_J$  for  $P_{\mathbb{R}^J}$ , the orthogonal projection onto the subspace  $\mathbb{R}^J$ . We denote the probability and expectation over the random subset  $J$  by  $\mathbb{P}_J$  and  $\mathbb{E}_J$ .

**Lemma 5.1.** *There exists constant  $c_2 > 0$  depending only on  $\rho$  such that for every  $x \in \text{Incomp}_n(\delta, \rho)$ , the event*

$$\mathcal{E}(x) := \left\{ \frac{P_J x}{\|P_J x\|_2} \in S^J \text{ and } \frac{\rho\sqrt{m}}{\sqrt{2n}} \leq \|P_J x\|_2 \leq \frac{\sqrt{m}}{\sqrt{\delta n}} \right\}$$

*satisfies  $\mathbb{P}_J(\mathcal{E}(x)) \geq (c_2 \delta)^m$ .*

*Proof.* Let  $\sigma \subset \{1, \dots, n\}$  be the subset from Lemma 3.4. Then we have

$$\mathbb{P}_J(J \subset \sigma) = \binom{|\sigma|}{m} / \binom{n}{m}.$$

Using Stirling's approximation, for  $d \leq \frac{1}{4} \rho^2 \delta n$ , we have

$$\mathbb{P}_J(J \subset \sigma) \geq \left( \frac{c|\sigma|}{n} \right)^m \geq (c_2 \delta)^m,$$

and for  $d > \frac{1}{4} \rho^2 \delta n$ , we have

$$\mathbb{P}_J(J \subset \sigma) \geq \binom{n}{m}^{-1} \geq \frac{m!}{n^m} \geq \left( \frac{cm}{n} \right)^m \geq (c_2 \delta)^m.$$

If  $J \subset \sigma$ , then summing (3.7) over  $k \in J$ , we obtain the required two-sided bound for  $\|P_J x\|_2$ . This and (3.7) yield  $P_J x / \|P_J x\|_2 \in S^J$ . Hence  $\mathcal{E}(x)$  holds.  $\square$

Lemma 5.1 implies the following lemma, whose proof is similar to the one for [17, Lemma 6.2].

**Lemma 5.2.** *Let  $J$  denote the  $m$ -element subsets of  $\{1, \dots, n\}$ . Then for every  $\epsilon > 0$ ,*

$$\mathbb{P} \left( \inf_{x \in \text{Incomp}_n(\delta, \rho)} \|Zx\|_2 < \epsilon \rho \sqrt{\frac{m}{2n}} \right) \leq (c_2 \delta)^{-m} \max_J \mathbb{P} \left( \inf_{x \in S^J} \text{dist}(Zx, H_{J^c}) < \epsilon \right). \quad (5.5)$$

It remains to bound  $\mathbb{P}(\inf_{x \in S^J} \text{dist}(Zx, H_{J^c}) < \epsilon)$  for any  $m$ -element subset  $J$ . We shall need the following lemma to bound below the distance between a random vector in  $\mathbb{R}^N$  and an independent random subspace of codimension  $l$ . It will be proved in Section 6.

**Lemma 5.3** (Distance to a random subspace). *Let  $J$  be any  $m$ -element subset of  $\{1, \dots, n\}$  and let  $H_{J^c}$  be the random subspace of  $\mathbb{R}^N$  defined in (5.3). Let  $X$  be a random vector in  $\mathbb{R}^M$  whose coordinates are i.i.d. centered random variables with unit variance and finite fourth moments, independent of  $H_{J^c}$ . Assume that  $l := m + d - 1 \leq \beta N$ . Then for every  $\epsilon > 0$ , we have*

$$\mathbb{P} \left( \sup_{v \in \mathbb{R}^N} \mathbb{P} \left( \text{dist}(PX - v, H_{J^c}) < \epsilon \sqrt{l} \mid H_{J^c} \right) > (\tilde{C}\epsilon)^l + e^{-\tilde{c}N} \right) \leq e^{-\tilde{c}N}, \quad (5.6)$$

where  $\beta, \tilde{c}, \tilde{C} > 0$  depend only on  $\Lambda, C_1$  and the maximal fourth moment.

It is easy to see that (5.6) implies the weaker result:

$$\sup_{v \in \mathbb{R}^N} \mathbb{P} \left( \text{dist}(PX - v, H_{J^c}) < \epsilon \sqrt{l} \right) \leq (\tilde{C}\epsilon)^l + 2e^{-\tilde{c}N}. \quad (5.7)$$

In the following proof, we choose  $\lambda_0$  such that

$$d \leq \beta N/2 \Rightarrow l \leq 2d \leq \beta N. \quad (5.8)$$

Note that for any fixed  $x \in S^J$ , we have  $Zx = PYx - Ax$ , where  $Yx$  is a random vector satisfying the assumptions for  $X$  in Lemma 5.3. So (5.7) gives a useful probability bound for a single  $x \in S^J$ . Then we will try to take a union bound over all  $x$  in an  $\epsilon$ -net of  $S^J$  and obtain a uniform distance bound. This is stated in the following theorem.

**Theorem 5.4** (Uniform distance bound). *Let  $Y$  be a random matrix satisfying the assumptions in Theorem 1.2. Then for every  $m$ -element subset  $J$  and  $t > 0$ ,*

$$\mathbb{P} \left( \inf_{x \in S^J} \text{dist}(Zx, H_{J^c}) < t\sqrt{d} \right) \leq (\tilde{C}tK^5 \log K)^d + e^{-\tilde{c}N}, \quad (5.9)$$

where  $\tilde{C}, \tilde{c} > 0$  depend only on  $\tilde{C}$  and  $\tilde{c}$ .

By the definition of  $m$  in (5.2), we have

$$(c_2\delta)^{-m} \leq \left[ (c_2\delta)^{-\rho^2\delta/2} \right]^n \leq e^{\tilde{c}N/2},$$

with an appropriate choice of  $\rho$ . Then we conclude from Lemma 5.2 and Theorem 5.4 that

$$\begin{aligned} \mathbb{P} \left( \inf_{x \in \text{Incomp}_n(\delta, \rho)} \|Zx\|_2 < \epsilon \rho \sqrt{\frac{md}{2n}} \right) &\leq (c_2\delta)^{-m} (\tilde{C}\epsilon K^5 \log K)^d + e^{-\tilde{c}N/2} \\ &\leq (CK^9(\log K)^2\epsilon)^d + e^{-\tilde{c}N/2}, \end{aligned} \quad (5.10)$$

where we used  $m \leq d$  and  $\delta$  in (5.1). Changing  $\epsilon$  to  $\epsilon \rho^{-1} \sqrt{2d/m}$  in (5.10) and using  $d/m \leq CK^4 \log K$ , we get

$$\mathbb{P} \left( \inf_{x \in \text{Incomp}_n(\delta, \rho)} \|(PY - A)x\|_2 < \epsilon \frac{d}{\sqrt{n}} \right) \leq \left( CK^{11}(\log K)^{5/2}\epsilon \right)^d + e^{-\tilde{c}N/2},$$

which, together with Lemma 3.2, concludes the proof of Theorem 1.2.

Now we begin the proof of Theorem 5.4. Without loss of generality, we can assume that the entries of  $Y$  have absolute continuous distributions. In fact we can add to each entry an independent

Gaussian random variable with small variance  $\sigma$ , and later let  $\sigma \rightarrow 0$  (all the estimates below do not depend on  $\sigma$ ). Under this assumption, we have the following convenient fact:

$$\dim(H_{J^c}) = n - m \quad \text{a.s.} \quad (5.11)$$

Let  $P_{H^\perp}$  be the orthogonal projection in  $\mathbb{R}^N$  onto  $H_{J^c}^\perp$ , and define

$$W := P_{H^\perp} P Y|_{\mathbb{R}^J}. \quad (5.12)$$

Then for every  $x \in \mathbb{R}^n$ , we have

$$\text{dist}(PYx - v, H_{J^c}) = \|Wx - w\|_2, \quad \text{where } w = P_{H^\perp} v. \quad (5.13)$$

By (5.11),  $\dim(H_{J^c}^\perp) = N - n + m = l$  almost surely. Thus  $W$  acts as an operator from an  $m$ -dimensional subspace into an  $l$ -dimensional subspace. If we have a proper operator bound for  $W$ , we can run the approximation argument on  $S^J$  and prove a uniform distance bound over all  $x \in S^J$ .

**Proposition 5.5.** *Let  $W$  be a random matrix as in (5.12). Then*

$$\mathbb{P}\left(\|W\| > sK\sqrt{d} \mid H_{J^c}\right) \leq e^{-c_0 s^2 d}, \quad \text{for } s \geq C_0,$$

where  $C_0, c_0 > 0$  are absolute constants.

*Proof.* For simplicity of notations, we fix a realization of  $H_{J^c}$  and omit the conditioning on it from the expressions below. Let  $\mathcal{N}$  be an  $(1/2)$ -net of  $S^{n-1} \cap \mathbb{R}^J$  and  $\mathcal{M}$  be an  $(1/2)$ -net of  $S^{n-1} \cap H_{J^c}^\perp$ . By Lemma 2.2, we can choose  $\mathcal{N}$  and  $\mathcal{M}$  such that

$$|\mathcal{N}| \leq 5^m, \quad |\mathcal{M}| \leq 5^l.$$

It is easy to prove that

$$\|W\| \leq 4 \sup_{x \in \mathcal{N}, y \in \mathcal{M}} |\langle Wx, y \rangle|. \quad (5.14)$$

For every  $x \in \mathcal{N}$  and  $y \in \mathcal{M}$ ,  $\langle Wx, y \rangle = \langle PYx, y \rangle = \langle Yx, P^T y \rangle$  is a random variable with subgaussian moment bounded by  $CK$  for some absolute constant  $C > 0$ . Hence by (1.13) we have

$$\mathbb{P}\left(|\langle Wx, y \rangle| > \frac{1}{4}sK\sqrt{d}\right) \leq 2e^{-cs^2 d}.$$

Using (5.14) and taking the union bound, we get that for large enough  $C_0$ ,

$$\mathbb{P}\left(\|W\| > sK\sqrt{d}\right) \leq 5^m \cdot 5^l \cdot 2e^{-cs^2 d} \leq e^{-c_0 s^2 d}, \quad \text{for } s \geq C_0,$$

where we used that  $m \leq l \leq 2d$ . □

**Lemma 5.6.** *Let  $W$  be a random matrix as in (5.12) and let  $w$  be a random vector as in (5.13). Then for every  $t \geq 0$ , we have*

$$\mathbb{P}\left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d}, \|W\| \leq C_0 K\sqrt{d}\right) \leq K^{m-1}(C_2 t)^d + 2e^{-\tilde{c}N/4}, \quad (5.15)$$

where  $C_2$  depends only on  $\tilde{C}$ .

*Proof.* Fix any  $x \in S^J$ . It is easy to verify that  $Yx$  is a random vector that satisfies the assumptions for  $X$  in Lemma 5.3. Hence by (5.13) and (5.7), we have

$$\mathbb{P}\left(\|Wx - w\|_2 < t\sqrt{d}\right) \leq \mathbb{P}\left(\text{dist}(PYx - v, H_{J^c}) < t\sqrt{l}\right) \leq (\tilde{C}t)^l + 2e^{-\tilde{c}N}. \quad (5.16)$$

Let  $\epsilon = t/(C_0K)$ . By Lemma 2.2, there exists an  $\epsilon$ -net  $\mathcal{N}$  of  $S^J$  with  $|\mathcal{N}| \leq 2m(5C_0K/t)^{m-1}$ . Consider the event

$$\mathcal{E}_t := \left\{ \inf_{x \in \mathcal{N}} \|Wx - w\|_2 < 2t\sqrt{d} \right\}.$$

Taking the union bound, we get that

$$\mathbb{P}(\mathcal{E}_t) \leq 2m \left( \frac{5C_0K}{t} \right)^{m-1} \left[ (2\tilde{C}t)^{m+d-1} + 2e^{-\tilde{c}N} \right] \leq K^{m-1}(C_2t)^d + 4m \left( \frac{5C_0K}{t} \right)^{m-1} e^{-\tilde{c}N}.$$

For  $t \geq t_0 := e^{-\tilde{c}N/(4d)}/(C_2K)$ , we have

$$4m \left( \frac{5C_0K}{t} \right)^{m-1} \leq (C'_0K^2)^{\rho^2\delta n/2} e^{\tilde{c}N/4} \leq e^{\tilde{c}N/2}$$

with an appropriate choice of  $\rho$ . Thus we get

$$\mathbb{P}(\mathcal{E}_t) \leq K^{m-1}(C_2t)^d + e^{-\tilde{c}N/2}, \quad \text{for } t \geq t_0.$$

For  $t < t_0$ , we have

$$\mathbb{P}(\mathcal{E}_t) \leq \mathbb{P}(\mathcal{E}_{t_0}) \leq K^{m-1}(C_2t_0)^d + e^{-\tilde{c}N/2} \leq 2e^{-\tilde{c}N/4}.$$

Then applying the standard approximation argument, we can check that the probability in (5.15) is bounded by  $\mathbb{P}(\mathcal{E}_t)$ , which concludes the proof.  $\square$

With Proposition 5.5 and Lemma 5.6, we obtain that

$$\mathbb{P}\left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d}\right) \leq K^{m-1}(C_2t)^d + 2e^{-\tilde{c}N/4} + e^{-c_0C_0^2d}.$$

Unfortunately, the bound  $e^{-c_0C_0^2d}$  is too weak for small  $d$ . Following the idea in [17], we refine the probability bound by decoupling the information about  $\|Wx - w\|_2$  from the information about  $\|W\|$ . The proof of next lemma is essentially the same as the one for Proposition 7.5 of [17]. We omit the details.

**Lemma 5.7** (Decoupling). *Let  $X$  be an  $N \times m$  matrix whose columns are independent random vectors, and let  $A$  be an  $N \times N$  deterministic matrix. Let  $z \in S^{m-1}$  be a vector satisfying  $|z_k| \geq K_1/\sqrt{m}$  for all  $k \in \{1, \dots, m\}$ . Then for every  $v \in \mathbb{R}^N$  and every  $0 < a < b$ , we have*

$$\mathbb{P}(\|AXz - Av\|_2 < a, \|AX\| > b) \leq 2 \sup_{y \in S^{m-1}, u \in \mathbb{R}^N} \mathbb{P}\left(\|AXy - Au\|_2 < \frac{\sqrt{2}a}{K_1}\right) \mathbb{P}\left(\|AX\| > \frac{b}{\sqrt{2}}\right).$$

**Remark 5.8.** By (5.4), all the vectors in  $S^J$  satisfy the assumption for  $z$  in Lemma 5.7.

With this decoupling lemma, we can prove the following refinement of Lemma 5.6.

**Lemma 5.9.** *Let  $W$  be a random matrix as in (5.12) and let  $w$  be a random vector as in (5.13). For every  $s \geq 1$  and every  $t \geq 0$ , we have*

$$\mathbb{P} \left( \inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d} \text{ and } sC_0K\sqrt{d} < \|W\| \leq 2sC_0K\sqrt{d} \right) \leq \left[ \frac{K^{m-1}(C_3t)^d}{K_1^{m+d-1}} + 2e^{-\tilde{c}N/4} \right] e^{-c_1s^2d},$$

where  $c_1$  is an absolute constant and  $C_3$  depends only on  $\tilde{C}$ .

*Proof.* Let  $\epsilon = t/(2sC_0K)$ . By Lemma 2.2, there exists an  $\epsilon$ -net  $\mathcal{N}$  of  $S^J$  with  $|\mathcal{N}| \leq 2m(9sC_0K/t)^{m-1}$ . Consider the event

$$\mathcal{E}_t := \left\{ \inf_{x \in \mathcal{N}} \|Wx - w\|_2 < 2t\sqrt{d} \text{ and } \|W\| > sC_0K\sqrt{d} \right\}.$$

Conditioning on  $H_{J^c}$ , we can apply Lemma 5.7 to get that

$$\mathbb{P}(\mathcal{E}_t | H_{J^c}) \leq |\mathcal{N}| \cdot 2 \sup_{x \in S^{m-1}, v \in \mathbb{R}^N} \mathbb{P} \left( \|Wx - P_{H^\perp} v\|_2 < \frac{\sqrt{2}}{K_1} \cdot 2t\sqrt{d} \mid H_{J^c} \right) \mathbb{P} \left( \|W\| \geq \frac{sC_0K\sqrt{d}}{\sqrt{2}} \mid H_{J^c} \right)$$

Taking expectation over  $H_{J^c}$  and using Proposition 5.5, we obtain that

$$\begin{aligned} \mathbb{P}(\mathcal{E}_t) &\leq 4m \left( \frac{9sC_0K}{t} \right)^{m-1} e^{-c_0C_0^2s^2d/2} \mathbb{E} \left[ \sup_{x \in S^{m-1}, v \in \mathbb{R}^N} \mathbb{P} \left( \|Wx - P_{H^\perp} v\|_2 < \frac{\sqrt{2}}{K_1} \cdot 2t\sqrt{d} \mid H_{J^c} \right) \right] \\ &\leq 4m \left( \frac{9C_0K}{t} \right)^{m-1} \left( s^{m-1} e^{-c_0C_0^2s^2d/2} \right) \left[ \left( \frac{2\sqrt{2}\tilde{C}t}{K_1} \right)^{m+d-1} + 2e^{-\tilde{c}N} \right], \end{aligned}$$

where in the second step we used the representation in (5.13) and the estimate (5.6). Since  $s \geq 1$  and  $1 \leq m \leq d$ , we can bound this as

$$\mathbb{P}(\mathcal{E}_t) \leq \left[ \frac{K^{m-1}(C_3t)^d}{K_1^{m+d-1}} + C_4m \left( \frac{9C_0K}{t} \right)^{m-1} e^{-\tilde{c}N} \right] e^{-c_1s^2d},$$

where  $C_4 > 0$  is an absolute constant. For  $t \geq t_1 := e^{-\tilde{c}N/(4d)} K_1^2/(C_3K)$ , we have

$$C_4m \left( \frac{9C_0K}{t} \right)^{m-1} \leq (C'_0K^2K_1^{-2})^{\rho^2\delta n/2} e^{\tilde{c}N/4} \leq e^{\tilde{c}N/2}$$

with an appropriate choice of  $\rho$ . Thus we get

$$\mathbb{P}(\mathcal{E}_t) \leq \left[ \frac{K^{m-1}(C_3t)^d}{K_1^{m+d-1}} + e^{-\tilde{c}N/2} \right] e^{-c_1s^2d}, \quad \text{for } t \geq t_1.$$

For  $t < t_1$ , we have

$$\mathbb{P}(\mathcal{E}_t) \leq \mathbb{P}(\mathcal{E}_{t_1}) \leq \left[ \frac{K^{m-1}(C_3t_1)^d}{K_1^{m+d-1}} + e^{-\tilde{c}N/2} \right] e^{-c_1s^2d} \leq 2e^{-\tilde{c}N/4} e^{-c_1s^2d}.$$

Suppose there exists  $y \in S^J$  such that

$$\|Wy - w\|_2 < t\sqrt{d} \quad \text{and} \quad sC_0K\sqrt{d} < \|W\| \leq 2sC_0K\sqrt{d}.$$

Then we choose  $x \in \mathcal{N}$  such that  $\|x - y\|_2 \leq \epsilon$ , and by triangle inequality we obtain that

$$\|Wx - w\|_2 \leq \|Wy - w\|_2 + \|W\|\|x - y\|_2 < t\sqrt{d} + 2sC_0K\sqrt{d}\epsilon \leq 2t\sqrt{d},$$

i.e. the event  $\mathcal{E}_t$  holds. Then the bound for  $P(\mathcal{E}_t)$  concludes the proof.  $\square$

*Proof of Theorem 5.4.* Summing the probability bounds in Lemma 5.6 and Lemma 5.9 for  $s = 2^k$ ,  $k \in \mathbb{Z}_+$ , we conclude that

$$\begin{aligned} \mathbb{P}\left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d}\right) &\leq K^{m-1}(C_2t)^d + 2e^{-\tilde{c}N/4} + \left[\frac{K^{m-1}(C_3t)^d}{K_1^{m+d-1}} + 2e^{-\tilde{c}N/4}\right] \sum_{s=2^k, k \in \mathbb{Z}_+} e^{-c_1 s^2 d} \\ &\leq (C_5 K K_1^{-2} t)^d + C_6 e^{-\tilde{c}N/4}. \end{aligned}$$

Using that  $K_1 = \rho\sqrt{\delta/2}$  (see (5.4)) and  $\delta = c_1 N / (nK^4 \log K)$  (see (5.1)), we get

$$\mathbb{P}\left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d}\right) \leq (CtK^5 \log K)^d + Ce^{-\tilde{c}N/4}.$$

In view of the representation (5.13), this concludes the proof.  $\square$

## 6 Proof of Lemma 5.3

We will first prove a general inequality that holds for any fixed subspace  $H$  in  $\mathbb{R}^N$  of codimension  $l = m + d - 1$ . This probability bound will depend on the arithmetic structure of  $H$ , which can be expressed using the *least common denominator* (LCD). Following the notations in [17], for  $\alpha > 0$  and  $\gamma \in (0, 1)$ , we define the least common denominator of a vector  $a \in \mathbb{R}^M$  as

$$\text{LCD}_{\alpha, \gamma}(a) := \inf \{ \theta > 0 : \text{dist}(\theta a, \mathbb{Z}^M) < \min(\gamma \|\theta a\|_2, \alpha) \}.$$

More generally, let  $a = (a_1, \dots, a_M)$  be a sequence of vectors  $a_k \in \mathbb{R}^l$ . We define the product of such multi-vector  $a$  and a vector  $\theta \in \mathbb{R}^l$  as

$$\theta \cdot a := (\langle \theta, a_1 \rangle, \dots, \langle \theta, a_M \rangle) \in \mathbb{R}^M.$$

Then we define, for  $\alpha > 0$  and  $\gamma \in (0, 1)$ ,

$$\text{LCD}_{\alpha, \gamma}(a) := \inf \{ \|\theta\|_2 : \theta \in \mathbb{R}^l, \text{dist}(\theta \cdot a, \mathbb{Z}^M) < \min(\gamma \|\theta \cdot a\|_2, \alpha) \}.$$

Finally, the least common denominator of a subspace  $E \subseteq \mathbb{R}^M$  is defined as

$$\text{LCD}_{\alpha, \gamma}(E) := \inf \{ \text{LCD}_{\alpha, \gamma}(a) : a \in S(E) \} = \inf \{ \|\theta\|_2 : \theta \in E, \text{dist}(\theta, \mathbb{Z}^M) < \min(\gamma \|\theta\|_2, \alpha) \}. \quad (6.1)$$

A key to the proof is the next small ball probability theorem.



**Theorem 6.1** (Theorem 3.3 of [17]). *Consider a sequence  $a = (a_1, \dots, a_M)$  of vectors  $a_k \in \mathbb{R}^l$ , which satisfies*

$$\sum_{k=1}^M \langle x, a_k \rangle^2 \geq \|x\|_2^2 \quad \text{for every } x \in \mathbb{R}^l. \quad (6.2)$$

*Let  $\xi_1, \dots, \xi_M$  be i.i.d. centered random variables, such that  $\mathcal{L}(\xi_k, 1) \leq 1 - b$  for some  $b > 0$ . Consider the random sum  $S := \sum_{k=1}^M a_k \xi_k \in \mathbb{R}^l$ . Then, for every  $\alpha > 0$  and  $\gamma \in (0, 1)$ , and for*

$$\epsilon \geq \frac{\sqrt{l}}{\text{LCD}_{\alpha, \gamma}(a)},$$

*we have*

$$\mathcal{L}(S, \epsilon \sqrt{l}) \leq \left( \frac{C\epsilon}{\gamma \sqrt{b}} \right)^l + C^l e^{-2b\alpha^2}.$$

Let  $H$  be a fixed subspace in  $\mathbb{R}^N$  of codimension  $l$ . We denote an orthonormal basis of  $H^\perp$  by  $\{n_1, \dots, n_l\} \subseteq \mathbb{R}^N$ , and write  $X$  in coordinates as  $X = (\xi_1, \dots, \xi_M)$ . Then using  $PP^T = 1$ , we get

$$\begin{aligned} \text{dist}(PX - v, H) &= \|P_{H^\perp}(PX - v)\|_2 = \left\| \sum_{r=1}^l \langle PX, n_r \rangle n_r - P_{H^\perp} v \right\|_2 = \left\| \sum_{r=1}^l \langle X, P^T n_r \rangle n_r - P_{H^\perp} v \right\|_2 \\ &= \left\| \sum_{r=1}^l \langle X, P^T n_r \rangle P^T n_r - P^T P_{H^\perp} v \right\|_2 = \|P_E X - w\|_2 = \left\| \sum_{k=1}^M a_k \xi_k - w \right\|_2, \end{aligned}$$

where

$$E \equiv E(H) := P^T H^\perp, \quad a_k := P_E e_k, \quad w := P^T P_{H^\perp} v,$$

and where  $e_1, \dots, e_M$  denote the canonical basis of  $\mathbb{R}^M$ . Notice that

$$\sum_{k=1}^M \langle x, a_k \rangle^2 = \|x\|_2^2, \quad \text{for any } x \in E.$$

Hence we can use Theorem 6.1 in the space  $E$  (identified with  $\mathbb{R}^l$  by a suitable isometry). For every  $\theta = (\theta_1, \dots, \theta_M) \in E$  and every  $k$ , we have  $\langle \theta, a_k \rangle = \langle \theta, e_k \rangle = \theta_k$ , so  $\theta \cdot a = \theta$ , where the right hand side is considered as a vector in  $\mathbb{R}^M$ . Therefore, we have

$$\text{LCD}_{\alpha, \gamma}(E) = \text{LCD}_{\alpha, \gamma}(a).$$

By Lemma 2.5,  $\mathcal{L}(\xi_k, 1/2) \leq 1 - b$  for some  $b > 0$  that depends only on the fourth moment of  $\xi_k$ . Hence we can apply Theorem 6.1 to  $S = \sum_{k=1}^M a_k \xi_k$  and conclude that for every  $\epsilon > 0$ ,

$$\mathbb{P}(\text{dist}(PX - v, H) < \epsilon \sqrt{l}) \leq \mathcal{L}(S, \epsilon \sqrt{l}) \leq \left( \frac{C\epsilon}{\gamma} \right)^l + \left( \frac{C\sqrt{l}}{\gamma \text{LCD}_{\alpha, \gamma}(E)} \right)^l + C^l e^{-c\alpha^2}. \quad (6.3)$$

Now it suffices to bound below the least common denominator of the random subspace  $E$ . Heuristically, the randomness should remove any arithmetic structure from the subspace  $E$  and make the LCD exponentially large. The next theorem shows that this is indeed true.

**Theorem 6.2.** Suppose  $\xi_1, \dots, \xi_{N-l}$  are independent centered random variables with unit variance and uniformly bounded fourth moment. Let  $\tilde{Y}$  be an  $M \times (N-l)$  random matrix whose rows are independent copies of the random vector  $(\xi_1, \dots, \xi_{N-l})$ , and  $\tilde{A}$  be an  $N \times (N-l)$  deterministic matrix. Suppose that  $\|\tilde{Y}\| + \|\tilde{A}\| \leq C_1\sqrt{N}$  for some constant  $C_1 > 0$ . Let  $H$  be the random subspace of  $\mathbb{R}^N$  spanned by the column vectors of  $P\tilde{Y} - \tilde{A}$ , and define the subspace  $E \equiv E(H) := P^T H^\perp \subseteq \mathbb{R}^M$ . Then for  $\alpha = c\sqrt{N}$ , we have

$$\mathbb{P}\left(\text{LCD}_{\alpha,c}(E) < c\sqrt{N}e^{cN/l}\right) \leq e^{-cN},$$

where  $c$  depends only on  $\Lambda$ ,  $C_1$  and the maximal fourth moment.

*Proof of Lemma 5.3.* Consider the event  $\mathcal{E} := \{\text{LCD}_{\alpha,c}(E(H_{J^c})) \geq c\sqrt{N}e^{cN/l}\}$ . The above theorem shows that  $\mathbb{P}(\mathcal{E}) \geq 1 - e^{-cN}$ . Conditioning on a realization of  $H_{J^c}$  in  $\mathcal{E}$ , we obtain from (6.3) that

$$\sup_{v \in \mathbb{R}^N} \mathbb{P}\left(\text{dist}(PX - v, H_{J^c}) < \epsilon\sqrt{l} \mid H_{J^c}\right) \leq (C'\epsilon)^l + (C')^l e^{-c'N}, \quad \text{for } H_{J^c} \in \mathcal{E}. \quad (6.4)$$

Since  $l \leq \beta N$ , with an appropriate choice of  $\beta$  we get

$$(C')^l \leq e^{c'N/2}.$$

Then the proof is completed by the estimate on the probability of  $\mathcal{E}^c$ .  $\square$

The rest of this section is devoted to proving Theorem 6.2. Note that if  $a \in E(H)$ , then  $a = P^T b$  for some  $b \in H^\perp$ . Then with  $b = Pa$ , we have that

$$b \in H^\perp \Leftrightarrow \tilde{Y}^T P^T b - \tilde{A}^T b = 0 \Leftrightarrow \tilde{Y}^T a - \tilde{A}^T Pa = 0.$$

We denote  $\tilde{B} := \tilde{A}^T P$ . For every set  $S$  in  $E$ , we have

$$\inf_{x \in S} \|\tilde{Y}^T x - \tilde{B}x\|_2 > 0 \text{ implies } S \cap E = \emptyset. \quad (6.5)$$

This helps us to “navigate” the random subspace  $E$  away from undesired sets  $S$  on the unit sphere.

As in Definition 3.1, we can define the compressible and incompressible vectors on  $S^{M-1}$ , which are denoted by  $\text{Comp}_M(\delta, \rho)$  and  $\text{Incomp}_M(\delta, \rho)$ , respectively. First, we have the following result for compressible vectors.

**Lemma 6.3** (Random subspaces are incompressible). *There exist  $\delta, \rho \in (0, 1)$  such that*

$$\mathbb{P}(E \cap \text{Comp}_M(\delta, \rho) = \emptyset) \geq 1 - e^{-c_0 N}, \quad (6.6)$$

where the constants  $\delta, \rho, c_0 > 0$  depend only on  $\Lambda$ ,  $C_1$  and the maximal fourth moment.

*Proof.* Due to (6.5), it suffices to prove that

$$\mathbb{P}\left(\inf_{x \in \text{Comp}_M(\delta, \rho)} \|\left(\tilde{Y}^T - \tilde{B}\right)x\|_2 \leq c_0\sqrt{N}\right) \leq e^{-c_0 N}. \quad (6.7)$$

In fact, the proof is similar to the one for Lemma 3.2. However, instead of Lemma 2.7, we will use the fact that  $\tilde{Y}^T$  has independent row vectors  $\tilde{Y}_1, \dots, \tilde{Y}_{N-l}$ . For any  $x \in S^{M-1}$ , it is easy to verify

that  $\langle \tilde{Y}_k, x \rangle$  has variance 1 and uniformly bounded fourth moment. Then by Lemma 2.5, there exists a  $p \in (0, 1)$  such that for any fixed  $v = (v_1, \dots, v_{N-l}) \in \mathbb{R}^{N-l}$ ,

$$\mathbb{P} \left( |\langle \tilde{Y}_k, x \rangle - v_k| \leq 1/2 \right) \leq p.$$

By Lemma 2.1, we can find constants  $\eta, \nu \in (0, 1)$  depending on  $p$  only and such that

$$\mathbb{P} \left\{ \|\tilde{Y}^T x - v\|_2 \leq \eta \sqrt{N-l} \right\} \leq \nu^{N-l}. \quad (6.8)$$

Recall that  $l \leq \beta N$  and  $M \leq \Lambda N$  by our assumptions. Then using (6.8) instead of (3.3), we can complete the proof of (6.7) as in Lemma 3.2.  $\square$

Fix the constants  $\delta$  and  $\rho$  given by Lemma 6.3 for the rest of this section. Note that in contrast to the case in Lemma 3.2,  $\delta$  is now an  $N$ -independent constant. We will further decompose  $\text{Incomp}_M(\delta, \rho)$  into level sets  $S_D$  according to the value  $D$  of the LCD. We shall prove a nontrivial lower bound on  $\inf_{x \in S_D} \|(\tilde{Y}^T - \tilde{B})x\|_2$  for each level set up to  $D$  of the exponential order. By (6.5), this means that  $E$  is disjoint from every such level set. Therefore,  $E$  must have exponentially large LCD. First, as a consequences of Lemma 3.4, we have the following lemma, which gives a weak lower bound for the LCD.

**Lemma 6.4** (Lemma 3.6 of [17]). *For every  $\delta, \rho \in (0, 1)$ , there exist  $c_1(\delta, \rho) > 0$  and  $c_2(\delta) > 0$  such that the following holds. Let  $a \in \text{Incomp}_M(\delta, \rho)$ . Then for every  $0 < c < c_1(\delta, \rho)$  and every  $\alpha > 0$ , one has*

$$\text{LCD}_{\alpha, c}(a) > c_2(\delta) \sqrt{M}.$$

**Definition 6.5** (Level sets). *Let  $D \geq c_2(\delta) \sqrt{M}$ . Define  $S_D \subseteq S^{M-1}$  as*

$$S_D := \{x \in \text{Incomp}_M(\delta, \rho) : D \leq \text{LCD}_{\alpha, c}(x) < 2D\} \cap (P^T \mathbb{R}^N).$$

To obtain a lower bound for  $\|(\tilde{Y}^T - \tilde{B})x\|_2$  on  $S_D$ , we use the  $\epsilon$ -net argument again. We first need such a bound for a single vector  $x$ . The proof of next lemma is very similar to the one for Lemma 4.6 in [17]. We omit the details.

**Lemma 6.6.** *Let  $x \in S_D$ . Then for every  $t > 0$  we have*

$$\mathbb{P} \left( \|(\tilde{Y}^T - \tilde{B})x\|_2 < t\sqrt{N} \right) \leq \left( Ct + \frac{C}{D} + Ce^{-c\alpha^2} \right)^{N-l}. \quad (6.9)$$

Now we construct a small  $\epsilon$ -net of  $S_D$ . Our argument here is a little harder than the one in [17], because the  $\epsilon$ -net lies in a subspace  $P^T \mathbb{R}^N \subseteq \mathbb{R}^M$ , whose direction is quite arbitrary. We shall need the following classical result in geometric functional analysis [4].

**Lemma 6.7.** *If  $S \subseteq \mathbb{R}^M$  is a subspace of codimension  $k$ , then*

$$|S \cap Q_M| \leq (\sqrt{2})^k,$$

where  $Q_M = [-1/2, 1/2]^M$  is the unit cube centered at the origin.

**Lemma 6.8.** *There exists a  $(4\alpha/D)$ -net of  $S_D$  of cardinality at most  $(CD/\sqrt{N})^N$ .*

*Proof.* We can assume that  $4\alpha/D \leq 1$ , otherwise the conclusion is trivial. For  $x \in S_D$ , we denote  $D(x) := \text{LCD}_{\alpha,c}(x)$ . By the definition of  $S_D$ , we have  $D \leq D(x) < 2D$ . By the definition of LCD, there exists  $p \in \mathbb{Z}^M$  such that

$$\|D(x)x - p\|_2 < \alpha. \quad (6.10)$$

Therefore,

$$\left\| x - \frac{p}{D(x)} \right\| < \frac{\alpha}{D(x)} \leq \frac{1}{4}.$$

Since  $\|x\|_2 = 1$ , it follows that

$$\left\| x - \frac{p}{\|p\|_2} \right\|_2 \leq \frac{2\alpha}{D}.$$

We can chose  $p$  such that it is the closest integer point to  $D(x)x$ . Since  $\|D(x)x\|_2 < 2D$ ,  $p$  must lie in the “cube covering”  $\tilde{F}$  of  $F := B(0, 2D) \cap P^T \mathbb{R}^N$ , defined as

$$\tilde{F} := \bigcup_{b \in F} \left( \prod_{i=1}^M [b_i - 1/2, b_i + 1/2] \right).$$

On the other hand, by (6.10) and using that  $\|D(x)x\|_2 < 2D$  and  $4\alpha/D \leq 1$ , we obtain

$$\|p\|_2 < D(x) + \alpha \leq 3D.$$

In sum, we get a  $(2\alpha/D)$ -net of  $S_D$  as:

$$\mathcal{N} := \left\{ \frac{p}{\|p\|_2} : p \in \mathbb{Z}^M \cap B(0, 3D) \cap \tilde{F} \right\}.$$

The cardinality of  $\mathcal{N}$  can be bounded by the volume of  $B(0, 3D) \cap \tilde{F}$ . By Fubini’s theorem, we have

$$|B(0, 3D) \cap \tilde{F}| \leq |B(0, 3D) \cap S| \cdot |S^\perp \cap Q_M|, \quad S := P^T \mathbb{R}^N.$$

Then using the volume formula for an  $N$ -dimension ball and Lemma 6.7, we obtain that

$$|\mathcal{N}| \leq (CD/\sqrt{N})^N.$$

Finally, we can find a  $4\alpha/D$ -net of the same cardinality, which lies in  $S_D$  (see Lemma 5.7 of [16]). This completes the proof.  $\square$

**Lemma 6.9.** *There exist  $c_3, c_4, \mu \in (0, 1)$  such that the following holds. Let  $\alpha = \mu\sqrt{N} \geq 1$  and  $D \leq c_3\sqrt{N}e^{c_3N/l}$ . Then*

$$\mathbb{P} \left( \inf_{x \in S_D} \left\| \left( \tilde{Y}^T - \tilde{B} \right) x \right\|_2 < c_4 N/D \right) \leq e^{-N}.$$

*Proof.* To conclude the proof, it is enough to find  $\nu > 0$  such that the event

$$\mathcal{E} := \left\{ \inf_{x \in S_D} \left\| \left( \tilde{Y}^T - \tilde{B} \right) x \right\|_2 < \frac{\nu N}{2D} \right\}$$

has probability  $\leq e^{-N}$ . Let  $\nu > 0$  be a small constant to be chosen later. We apply Lemma 6.6 with  $t = \nu\sqrt{N}/D$ . By the assumptions on  $\alpha$  and  $D$ , the term  $Ct$  dominates in the right hand side of (6.9). This gives for arbitrary  $x \in S_D$ ,

$$\mathbb{P}\left(\left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 < \frac{\nu N}{D}\right) \leq \left(\frac{C\nu\sqrt{N}}{D}\right)^{N-l}.$$

We take the  $(4\alpha/D)$ -net  $\mathcal{N}$  of  $S_D$  given by Lemma 6.8, and take the union bound to get

$$p := \mathbb{P}\left(\inf_{x \in \mathcal{N}} \|(Y^T - B)x\|_2 < \frac{\nu N}{D}\right) \leq \left(\frac{CD}{\sqrt{N}}\right)^N \left(\frac{C\nu\sqrt{N}}{D}\right)^{N-l} \leq \left(\frac{CD}{\sqrt{N}}\right)^l (C'\nu)^{N-l}.$$

Using the assumption on  $D$ , we can choose  $\nu$  small enough such that

$$p \leq (C'')^l e^{c_3 N} (C'\nu)^{N-l} \leq e^{-N},$$

where we used  $l \leq \beta N$  in the last step.

Now assume  $\mathcal{E}$  holds. By the assumption of Theorem 6.2, we have

$$\|\tilde{Y}^T - \tilde{B}\| \leq \|\tilde{Y}\| + \|\tilde{A}\| \leq C_1\sqrt{N}.$$

Fix  $x \in S_D$  such that  $\|(\tilde{Y}^T - \tilde{B})x\|_2 < \nu N/(2D)$ . Then we can find  $y \in \mathcal{N}$  such that  $\|x - y\| \leq 4\alpha/D$ . Then, by the triangle inequality we have

$$\left\|\left(\tilde{Y}^T - \tilde{B}\right)y\right\|_2 \leq \left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 + \left\|\tilde{Y}^T - \tilde{B}\right\| \cdot \|x - y\|_2 \leq \frac{\nu N}{2D} + C_1\sqrt{N} \frac{4\mu\sqrt{N}}{D} < \frac{\nu N}{D},$$

if we choose  $\mu < \nu/(8C_1)$ . Thus we get

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}\left(\inf_{x \in \mathcal{N}} \|(Y^T - B)x\|_2 < \frac{\nu N}{D}\right) \leq e^{-N},$$

which concludes the proof.  $\square$

*Proof of Theorem 6.2.* Consider  $x \in S^{M-1} \cap E$  such that

$$\text{LCD}_{\alpha,c}(x) < c_3\sqrt{N}e^{c_3 N/l}.$$

Then, by Lemma 6.4 and Definition 6.5, either  $x$  is compressible or  $x \in S_D$  for some  $D \in \mathcal{D}$ , where

$$\mathcal{D} := \left\{D : c_2\sqrt{N} \leq D < c_3\sqrt{N}e^{c_3 N/l}, D = 2^k, k \in \mathbb{N}\right\},$$

where we used that  $M \geq N$ . Therefore, we can decompose the desired probability as follows:

$$p := \mathbb{P}\left(\text{LCD}_{\alpha,c}(E) < c_3\sqrt{N}e^{c_3 N/l}\right) \leq \mathbb{P}(E \cap \text{Comp}_M(\delta, \rho) \neq \emptyset) + \sum_{D \in \mathcal{D}} \mathbb{P}(E \cap S_D \neq \emptyset).$$

The first term can be bounded by  $e^{-c_0 N}$  by Lemma 6.3. The other terms can be bounded with (6.5) and Lemma 6.9:

$$\mathbb{P}(E \cap S_D \neq \emptyset) \leq \mathbb{P}\left(\inf_{x \in S_D} \left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 = 0\right) \leq e^{-N}.$$

Since there are  $|\mathcal{D}| \leq CN$  terms in the sum, we conclude that

$$p \leq e^{-c_0 N} + CN e^{-N} \leq e^{-c' N}.$$

This concludes the proof.  $\square$

## References

- [1] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.
- [2] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, volume 2 of *Mathematics Monograph Series*. Science Press, Beijing, 2006.
- [3] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993.
- [4] K. Ball. *Volumes of sections of cubes and related problems*, pages 251–260. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.
- [5] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.
- [6] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and banach spaces. volume 1 of *Handbook of the Geometry of Banach Spaces*, pages 317 – 366. North-Holland, Amsterdam, 2001.
- [7] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
- [8] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.*, 6(3):440–449, 1965.
- [9] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.
- [10] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *arXiv:1410.3516*.
- [11] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491 – 523, 2005.
- [12] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics*, 195(2):491 – 523, 2005.
- [13] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.
- [14] G. Pan and W. Zhou. Circular law, extreme singular values and potential theory. *J. Multivar. Anal.*, 101(3):645–656, 2010.
- [15] M. Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. Math.*, 168(2):575–600, 2008.
- [16] M. Rudelson and R. Vershynin. The littlewood-offord problem and invertibility of random matrices. *Adv. Math.*, 218:600–633, 2008.
- [17] M. Rudelson and R. Vershynin. The smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62:1707–1739, 2009.
- [18] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:9 pp., 2013.

- [19] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Notices*, 2015(19):9594–9617, 2015.
- [20] J. W. Silverstein. The smallest eigenvalue of a large dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.
- [21] T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.
- [22] T. Tao, V. Vu, and M. Krishnapur. Random matrices: Universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010.
- [23] H. Xi, F. Yang, and J. Yin. Local circular law for the product of a deterministic matrix with a random matrix. *arXiv:1603.04066*.
- [24] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4):509–521, 1988.